International Journal of scientific Research & Engineering Technology (IJSET)
Vol.20pp. 68-75

SimAT: a new measure of semantic similarity

Abdoulaye Diallo[#1], Mouhamadou Thiam[#2], Moussa Lo[*3]

[#]*Iba Der Thiam University of Thiès*

[1]abdoulaye.diallo4@univ-thies.sn

[2]mthiam@univ-thies.sn

[*]*Gaston Berger University*

[3]moussa.lo@ugb.edu.sn

***Abstract*— Measures of similarity are used in the context of ontological reconciliation. There are several types of semantic similarity measures. Semantic similarity measures are among the most commonly used measures in this reconciliation. Some of these semantic measures exploit information content, some of them exploit depth, while others exploit the number of concept instances in the ontology. These measures therefore exploit only part of the information contained in the concepts. In this paper, we propose a new measure of semantic similarity, *SimAT,* based on informational content and concepts depth. The main originality of our measurement is that it exploits more information contained in a concept. The experimentation of our measurement was done on a hierarchy of concepts taken from WordNet. The calculation of the information content of a given concept is based on the probability of finding an instance of the concept. Our comparative study showed that our measurement performed better compared to measures based on depth or information content.**

***Keywords*— ontological reconciliation, semantic similarity, depth, informational content**

I. INTRODUCTION

The notion of similarity is widely used in several fields such as reconciliation or ontological alignment. Alignment approaches are a set of actions to achieve this reconciliation. Semantic similarity measures have been the topic of several scientific publications due to their usefulness in the context of this reconciliation. They are often used in tasks such as semantic grouping, information retrieval, lexical disambiguation and other areas of natural language processing.

There are several measures of semantic similarity that can be classified into different categories [1]:

- Structural-type measures using ontological structure and relationships between concepts to calculate similarity. For example, we can cite the measurements of Rada et al [2], Resnik [3], Leacock & Chodorow [4], Wu & Palmer [5], etc.
- Intentional-type measurements that are mainly based on the concepts to be matched. This is the example of Tversky's measure [6].
- Extensional measures are based on instances of two concepts to reconcile, and these concepts can have multiple instances. This is the case of the measures of Jaccard [7], Dice [8], D'Amato et al [9], etc.
- Expressional measures that are interested in the different terms that denote concepts to be matched, such as measures of Lin [10], Jian & Conrath [11], etc.

All of the presented measures above are essentially based either on the depth of the ontology, is the information content or on the number of instances of any concept. Therefore, these measures exploit only a part of the information contained in the concepts.

In this paper, we present *SimAT* similarity measure that allows us to exploit several facets of the information contained in a concept in order to improve the calculation of similarity. This new measurement has the particularity of improving Lin's measurement, which is based on information content as well as Wu & Palmer's measurement, which is based on depth. In our approach, information about the concepts depth in the ontology is combined with information content as shown in fig.1.

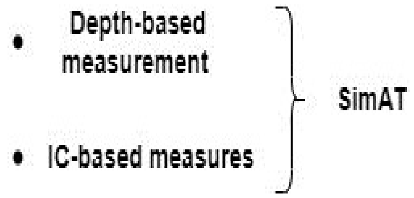

Fig 1 SimAT features

In the following, our document is structured as follows. Section 2 presents different measures of semantic similarity based on information content and depth. Section 3 present the details of our measure *SimAT*. Section 4 presents the results of our experimentation and the comparison of our measurement with other measures of semantic similarity.

II. RELATED WORK

There are many measures of semantic similarity. Calculating similarity between two concepts allows to determine the existing relationships between them. These measures include those based on depth, on information content and on instances. In this paper, we focus on measures based on depth and information content.

ISSN: 2356-5608

A. Depth-based measurements

Depth-based measurements are part of arc-based measurements [12]. To calculate similarity, they are based on the structuring of the hierarchy whose nodes are the concepts and the arcs are the links between the nodes or concepts. Many of these measures are presented below.

1) Rada's Measure

Rada et al [2] were the first to make the following hypothesis: the similarity between two concepts in a semantic network can be computed based on taxonomic links of the type <<*is-a*>>. The Rada measure is defined as the distance between the two concepts to be matched. This measure uses the distance that corresponds to the shortest path between concepts. It is given by the following expression.

$$Sim_{Rada}(c_1,c_2) = \frac{1}{1 + distance(c_1,c_2)}$$

Where $distance$ (c_1, c_2) is the number of arcs between the two concepts.

2) *Leacock and Chodorow measure*

The Leacock and Chodorow measure [4] are based on the Rada et al measure [2]. It uses the distance between concepts in a hierarchy and a logarithm to normalize the measurement to the maximum depth of the entire hierarchy. The formula is given by:

$$Sim_{Lea}(c_1,c_2) = -\log\left(\frac{dist_{edge}(c_1,c_2)}{2 * \max\_depth}\right)$$

where $dist_{edge}$(c_1, c_2) is the length of the shortest path between the two nodes and *max_depth* the maximum depth of the hierarchy.

3) *Wu & Palmer measure*

The Wu & Palmer measure [5] is one of the arc-based similarity measures. It uses the depth of concepts in ontology and measures similarity by taking into account the distance between concepts and their closest common ancestor (lcs). The formula for calculating the Wu & Palmer similarity between two concepts C_1 and C_2 is as follows:

$$Sim_{WuP}\ (c_1,c_2) = 2 * \frac{depth(lcs(c_1,c_2))}{(depth(c_1) + depth(c_2))}$$

Pekar & Staab measure

Pekar & Staab's measure [13] uses the distance between concepts and their smallest common ancestor (lcs) to calculate similarity.

$$Sim_{PeSt}(c_1,c_2) = \frac{depth(lcs)}{length(C_1,lcs) + length(C_2,lcs) + depth(lcs)}$$

Where *depth(lcs)* is the depth of the smallest common ancestor and *length (C_i, lcs)* is the distance between a concept and the smallest common ancestor.

4) *Nguyen & Al-Mubaid measure*

Nguyen & Al-Mubaid's measure [14] uses the maximum depth of the hierarchy and the smallest common ancestor (lcs) to calculate similarity between concepts. The formula used by the measure is:

$$Sim_{NgAl}(c_1,c_2) = \log\left(2 + (length(C_1,C_2) - 1) \times \left(depth(L) - depth(lcs)\right)\right)$$

Where L is the maximum depth of the hierarchy, *depth* and *length* are as defined in the Pekar & Staab measure.

5) *Zargayouna measure*

The Zargayouna measure [15] is an extension of the Wu & Palmer [5] measure. Like this latter, it exploits, in addition to depth, the distance between the concepts to be matched and the lowest concept of the taxonomy through a new parameter called specificity. The latter is denoted *Spec (C_1, C_2)* and represents the number of arcs that separate the two concepts at the lowest concept.

$$Sim_{sto}(c_1,c_2) = \frac{P_{ppcg} + 1}{(P_1 + 1) + (P_2 + 1) - (P_{ppcg} + 1)}$$

Where P_{ppcg} is the depth of the smallest generalizing concept and P_i is the depth of the concept C_i.

6) *Zhong measure*

Zhong's measure of similarity [17] is based on Zhong's distance. The latter is calculated using the smallest generalizing concept *ppcg*.

$$dist\,(c_1,c_2) = \frac{1}{2^{P_{ppcg}}} - \frac{1}{2^{P_1+1}} - \frac{1}{2^{P_2+1}}$$

The formula for Zhong similarity is as follows:

$$Sim_{Zhong}(c_1\,,c_2) = 1 - dist(c_1,c_2)$$

B. Information-content based measures

Measures based on information content can be classified among node-based measures [18]. To measure semantic similarities based on information content (IC), information content is used as a measure of the information contained in the concept. To calculate it, we associate a probability of occurrence for each class or instance in the ontology and the number of occurrences of these classes or instances [19].
Some of these measures are discussed in the following sections.

1) *Resnik's measure*

Resnik [3] introduced the simplest measure of semantic similarity based on information content (IC). It uses the amount of information shared by two concepts to measure their similarity. The measure proposed by Resnik has been widely used as a reference for comparing new measures of semantic similarity based on informational content.

ISSN: 2356-5608

International Journal of scientific Research & Engineering Technology (IJSET)
Vol.20pp. 68-75

$$Sim\ Res\ (c_1, c_2) = CI(ppg(c_1, c_2))$$

With *ppg* being the smallest generalizing concept.

2) Lin measure

Lin [10] proposed a measure of semantic similarity based on information content that improves Resnik's one by taking into account the amount of information contained in each concept. This measure is very often used in the calculation of semantic similarity based on information content. It uses the information content of the smallest common parent (C_{com}) in the calculation of similarity.

$$Sim_{Lin}(c_1, c_2) = \frac{2 * \psi(C_{com})}{\psi(c1) + \psi(c2)}$$

$\psi(Ci)$ represents the information content of the concept C_i.

3) Jian & Conrath measure

Jian & Conrath's measure [11] is based on the informational content (IC) of concepts in an ontology and measures similarity by taking into account the semantic distance between concepts. It also uses the information content of the smallest common parent between concepts. The formula for calculating the Jiang & Conrath similarity between two concepts C_1 and C_2 is as follows:

$$Sim_{JCn}(c_1, c_2) = \frac{1}{dist\ jcn}$$

$$dist_{\ jcn}\ (c_1, c_2) = IC(c_1) + IC(c_2) - 2 * (IC(C))$$

Where *C* is the smallest generalizing concept.

4) Faith measure

The Faith measure [20] is also part of the semantic similarity measures based on information content. It is equal to the information content of the most specific common concept (lso) divided by the difference between the sum of the information content of these concepts minus the informational content of the most specific common concept.

$$Sim_{FaITH}(c_1, c_2) = \frac{IC(lso(c_1, c_2)}{IC(c_1) + IC(c_2) - IC(lso(c_1, c_2)}$$

There are other measures based on informational content that combine information content and the representation of concepts (e.g., Rdf representation) for the calculation of similarity [21].

III. Presentation of Our Measure

Ontologies are defined as an explicit specification of a conceptualization [22]. They are composed of concepts organized according to well-defined hierarchy. Semantic similarity measures are used to calculate the similarity between two ontology concepts. Each concept occupies a level that is defined as its depth denoted *depth(c)*. Concepts within the hierarchy have *IC(c)*, information content. The information content of a concept in an ontology can be defined as the amount of information it provides about the entire ontology. Among others, it can be used to measure the specificity or generality of a concept. The more specific a concept is, the higher its information content. It is usually computed using information theory and the probability of occurrences of terms throughout the ontology. The informational content is calculated with the following formula [23]:

$$IC(c) = -\log(P(c))$$

where *P(c)* is the probability of finding an instance of concept *C*. These probabilities are calculated using the following formula:

$$P(c) = frequency(C) / N$$

Where *N* is the total number of concepts.

There are other formulas that can be used to calculate the information content of a concept using hyponyms. Some of these are presented in the following paragraphs.

- Seco & al

Seco et al [24] were the first to propose the computation of information content based on the number of hyponyms of the concept. Since *Hypo(c)* is the number of hyponyms in the taxonomic tree under concept C and *max_nodes* is the maximum number of concepts in the taxonomy, the proposed method for calculating the information content is as follows:

$$Ic_{seco\&al}(c) = 1 - \frac{(\log(hypo(c) + 1))}{\log(\max\ \_nodes)}$$

The denominator (corresponding to the most informative concept: c leaf in the tree) produces normalized information content values in the range of 0 to 1.

One of the limitations of this approach is that it only takes into account the hyponyms of a given concept in the taxonomy. Thus, concepts with the same number of hyponyms but different degrees of generality (i.e., one appears at a higher level of the hierarchy relative to the other) will therefore also be similar. This problem has been solved in the computational approach proposed by Zhou et al [24].

- Zhou & al

Zhou et al [24] proposed to complement the computation of hyponym-based information content with the relative depth of the concept in the taxonomy.

$$Ic_{Zhou\ \&\ al}(c) = k \times (1 - \frac{\log(hypo(C) + 1)}{\log(\max\ \_nodes)}) + (1 - k) \times (\frac{\log(depth(C))}{\log(\max\ \_depth)})$$

In addition to *hypo* and *max_nodes*, which have the same meaning as in the equation proposed by Seco et al, *depth(c)* corresponds to the depth of concept *c* in the taxonomy and *max_depth* is the maximum depth of the taxonomy. *k* is a tuning factor that adjusts the weight of the two characteristics involved in the evaluation of information content. The various

ISSN: 2356-5608

International Journal of scientific Research & Engineering Technology (IJSET)
Vol.20pp. 68-75

classical models used to calculate the IC do not use the number of occurrences of the concepts that subsume it. However, there are other methods that consider in the calculation of the IC the number of occurrences of the concepts that subsume the concept [25].

Our approach focuses on depth and information content in the calculation of similarity. Each *C* concept has a *depth(C)* associated with it. We base this on the fact that within the hierarchy, the further apart the concepts are, the less similar they are.

The results of the application of the Wu & Palmer [5] measurement applied to Fig. 3 are presented in Fig 2. These results show that the further apart the concepts are within the hierarchy, the less similar they are.

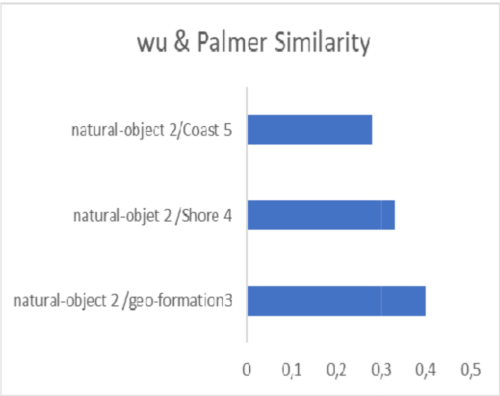

Fig.2 Similarity Wu & Palmer

When calculating the similarity between two concepts that have different depths, the result depends on the greater one between the two concepts. The greater the depth is, the less similar the two concepts are. So, in the *SimAT* measure we multiply by two the greatest depth between the two concepts and the result is divided by the sum of the depths of the two concepts.

Our calculation formula is made up of two factors:

$$Sim_{AT}(C_1, C_2) = \underbrace{\frac{2*\psi(C_{com})}{\psi(C_1)+\psi(C_2)}}_{\textbf{Factor 1}} * \underbrace{\frac{2*depth_{max}(C_1, C_2)}{depth(C_1)+depth(C_2)}}_{\textbf{Factor 2}}$$

Factor 1 corresponds to the Lin measure presented above based on information content. *Factor 2* represents the part concerning the depth of the concepts to be aligned. In our work, we exploit information about the depth of concepts within the hierarchy.

Together, these two factors allow for better use of information contained in the concepts.

IV. Experimentation

To experiment with our measurement, we used a hierarchy of concepts. This hierarchy is taken from Wordnet with the probabilities of the concepts. In this hierarchy, we note that each time the deeper the concept, the less likely it is and the higher its information content. These probabilities are shown in Fig. 3.

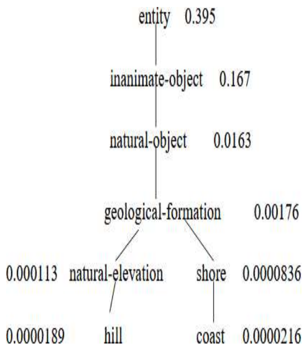

Fig. 3 wordnet extract

Using the formula presented in the section 3 to calculate the information content (eq. IC(c) = -log(P(c))) we obtained the results reported in Fig. 4. To facilitate the use of concepts, we have also added the depth of each one.

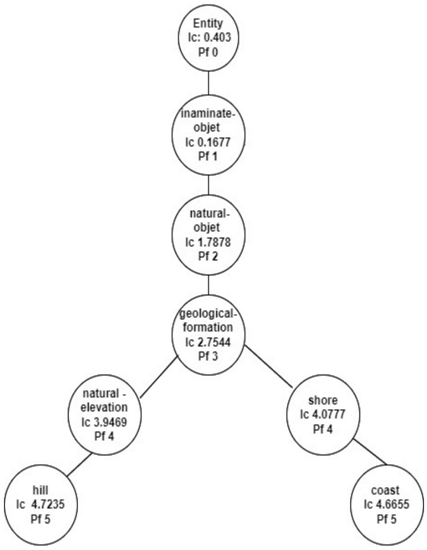

Fig. 4 wordnet extract with IC

Similarities between concepts were calculated and the results are reported in Table 1.

ISSN: 2356-5608

TABLE 1
COMPARISON OF LIN AND SIMAT

Concept & Depth	Lin measure	SimAT measure	Difference
nat-obj:2, geol-form:3	0.339956	0.407947	0.067991
geol_form:3, nat-elevation:4	0.531661	0.607612	0.075952
Shore:4, coast:5	0.696908	0.774343	0.077434
geo-form:3, coast:5	0.531661	0.664576	0.132915
nat-obj:2, coast:5	0.268949	0.384212	0.115264
nat-obj:2, shore:4	0.268949	0.358598	0.089650
inam-obj:1, coast:5	0.148431	0.247385	0.098954

In Table 1, we notice that our measure *SimAT* improves the Lin measure when calculating the similarity between two concepts that do not have the same depth within the hierarchy. The difference between the Lin measure and our measurement is proportional to the difference between the two depths of the concepts.
The results of the comparison between *SimAT*, Lin's measurement, and Wu & Palmer [5] are shown in Fig 5.

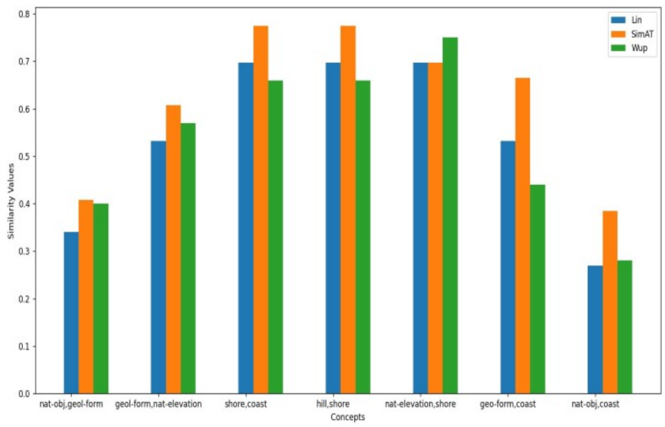

Fig.5 Comparison of Lin, Wu & Palmer and SimAT measurements

The histogram in Fig. 5 shows the results obtained with *SimAT* compared to the measurements of Lin and Wu & Palmer [5]. The concepts to be matched were chosen according to different depths. We find that our measurement improves the calculation in cases where the two concepts have different depths. This improvement grows according to the difference between the depths of the concepts.
We also calculated the similarity between the following concepts using depth-based measurements such as Rada [2], Wu & Palmer [5], Stojanovic for a better comparison with our measurement. The results are presented in Table 2.

TABLE 2
COMPARISON BETWEEN SIMAT AND DEPTH-BASED MEASURES

Concept: depth	wup	Rada	stoj	SimAT
nat-obj:2, geol-form:3	0.4	0.5	0.4	0.407947
geol_form:3, nat-elevation:4	0.5714	0.5	0.5	0.607612
Shore:4, coast:5	0.6666	0.5	0.5714	0.774343
geo-form:3, coast:5	0.5	0.33	0.4285	0.664576
nat-obj:2, coast:5	0.2857	0.25	0.2857	0.384212
nat-obj:2, shore:4	0.6666	0.33	0.3333	0.358598

In Table 2 we note that our measure *SimAT* performs better in most cases when compared to depth-based similarity measures (Wu & Palmer, Rada, Stojanovic).
To better interpret the results, we have also compared our approach with measures based on information content. The results are presented in Table 3.

TABLE 3
COMPARISON BETWEEN SIMAT AND IC-BASED MEASURES

Concept: depth	Lin	Faith	Jian&C	SimAT
nat-obj:2, geol-form:3	0.3399	0.5	0.2377	0.407947
geol_form:3, nat-elevation:4	0.5316	0.5	0.3199	0.607612
Shore:4, coast:5	0.6969	0.5	0.3091	0.774343
geo-form:3, coast:5	0.5316	0.33	0.2601	0.664576
nat-obj:2, coast:5	0.2689	0.25	0.1634	0.384212
nat-obj:2, shore:4	0.2689	0.33	0.1808	0.358598

Table 3 shows that the *SimAT* measure gives better results compared to the Lin measure (than the Faith and Jian & Conrath measures, respectively). We can notice that from the first calculation of the similarity between the concepts natural*_object* and *geological_formation* (for the Faith measure) the *SimAT* measure gives better results compared to other measures.
Table 4 summarizes the hole experimentation. The results obtained with all measures are presented in the table 4.

ISSN: 2356-5608

TABLE 4
COMPARISON BETWEEN SIMAT AND OTHER MEASURE

Concept: depth	Wup	Rada	Sto	Lin	Faith	J&C	SimAT
nat-obj:2, geol-form:3	0.4	0.5	0.4	0.33	0.5	0.23	0.40
geol_form:3, nat-elevation:4	0.57	0.5	0.5	0.53	0.5	0.319	0.60
Shore:4, coast:5	0.66	0.5	0.57	0.69	0.5	0.30	0.77
geo-form:3, coast:5	0.5	0.33	0.42	0.53	0.33	0.26	0.66
nat-obj:2, coast:5	0.28	0.25	0.28	0.26	0.25	0.16	0.38
nat-obj:2, shore:4	0.66	0.33	0.33	0.26	0.33	0.18	0.35

These results are emphasized in fig .6.
Fig. 6 shows the results of *SimAT* measurement and other based on depth and information content for better exploitation.

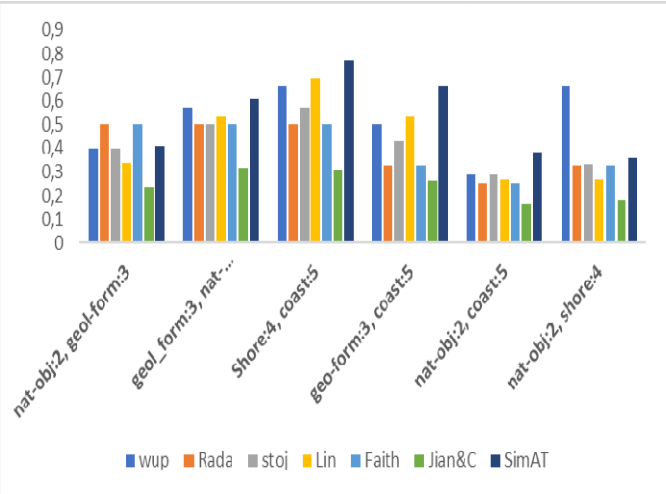

Fig.6 Comparison of SimAT with other measures

The histogram in Fig.6 shows that the *SimAT* measurement provides satisfactory results compared to measurements based on information content and depth. The *SimAT* measurement gives better results in 66% of cases compared to our tests on the calculation of the similarity between the concepts used in the experiment.

V. CONCLUSIONS AND PERSPECTIVES

Similarity measures are widely used in ontology alignment. The choice of the similarity measure depends on the nature of the data, the context of application and the specific requirements of the given task. Semantic similarity measures, on the other hand, use either information content, depth or instances for the calculation of similarity. Our new measure *SimAT* combines depth and informational content for the calculation of similarity. The parameter added to the Lin measurement is depth. In the calculation, we choose the greatest depth between the two concepts to align because the result depends on the difference between the two depths. Experimentation shows that our measurement improves the Lin measurement and the Wu & Palmer one for all concepts that do not have the same depth.
The *SimAT* similarity measure works best when the difference between the concepts to be matched is larger.
However, there are some limitations to our measure. This is the case, for example, where the two concepts to be aligned are at the same level, i.e., have the same depth. In this case, it's just a matter of calculating Lin's similarity.
These semantic measures presented above only calculate the similarity between two concepts and not between a concept and a sentence or between two sentences. There are, however, some measures that can be used to calculate the similarity between two sentences or between a word and a sentence using meaning, synonyms, hyperonyms, and homonyms [27]. Other semantic measures also exploit the labelling of semantic roles between sentences to calculate similarity [28].
The objective of our future work is to propose a new measure that will capitalize on more information contained in a concept, namely depth, informational content and number of instances.

ACKNOWLEDGMENT

This work was fully funded by the PASET/RSIF (Partnership for Skills in Applied Sciences, Engineering & Technologies/Regional Scholarship and Innovation Fund) through a doctoral scholarship. We would also like to thank CEAMITIC (African Centre of Excellence in Mathematics, Computation and ICT) for their support in our work.

REFERENCES

[1] Aimé, X. (2011). *Gradients de prototypicalité, mesures de similarité et de proximité sémantique : une contribution à l'Ingénierie des Ontologies (*Doctoral dissertation, Université de Nantes) p.82.
[2] Rada, R., et al. (1989) *Development and Application of a Metric on Semantic Nets*. IEEE Transactions on Systems, Man,andCybernetics,17-30. https://doi.org/10.1109/21.24528
[3] Resnik, P. (1995). *Using information content to evaluate semantic similarity in a taxonomy*. arXiv preprint cmp-lg/9511007.
[4] Leacock C., Chodorow M., « *Combining local context and WordNet sense similarity for word sense identification* », In WordNet, An Electronic Lexical Database. The MIT Press, 1998.
[5] Palmer, M., & Wu, Z. (1995). *Verb semantics for English-Chinese translation*. Machine translation, 10, 59-92.
[6] Tversky, Amos. "Features of similarity." *Psychological review* 84.4 (1977): 327.352
[7] P. Jaccard, “Distribution de la flore alpine dans le bassin des Dranses et dans quelques régions voisines”, Bulletin de la Société Vaudoise des Sciences Naturelles, vol. 37, pp. 241–272, 1901.

[8] Diep, H. N. (2002). Similarité de mots et extraction automatique de synonymes. *University of Louvain. Internship Report, Belgium*.

[9] d'Amato, C., Staab, S., & Fanizzi, N. (2008, September). On the influence of description logics ontologies on conceptual similarity. In *International Conference on Knowledge Engineering and Knowledge Management* (pp. 48-63). Berlin, Heidelberg: Springer Berlin Heidelberg.

[10] Lin, D. (1998, July). *An information-theoretic definition of similarity*. In *Icml* (Vol. 98, No. 1998, pp. 296-304).

[11] Jiang, J. J., & Conrath, D. W. (1997). *Semantic similarity based on corpus statistics and lexical taxonomy. arXiv preprint cmp-lg/9709008*.

[12] Ngom, A. N. (2015). Étude des mesures de similarité sémantique basées sur les arcs. In *CORIA* (pp. 535-544).

[13] Viktor Pekar and Steffen Staab. 2003. Word classification based on combined measures of distributional and semantic similarity. In *10th Conference of the European Chapter of the Association for Computational Linguistics*, Budapest, Hungary. Association for Computational Linguistics.

[14] Al-Mubaid H, Nguyen HA. A cluster-based approach for semantic similarity in the biomedical domain. Conf Proc IEEE Eng Med Biol Soc. 2006; 2006:2713-7. doi: 10.1109/IEMBS.2006.259235. PMID: 17946134.

[15] Zargayouna, H., & Salotti, S. (2004, May). *Mesure de similarité dans une ontologie pour l'indexation sémantique de documents XML*. In 15èmes Journées francophones d'Ingénierie des Connaissances (pp. 249-260). Presses universitaires de Grenoble.

[16] Stojanovic, N., Maedche, A., Staab, S., Studer, R., & Sure, Y. (2001, October). *SEAL: a framework for developing SEmantic PortALs.* In Proceedings of the 1st international conference on Knowledge capture (pp. 155-162).

[17] Zhong, J., Zhu, H., Li, J., & Yu, Y. (2002). Conceptual graph matching for semantic search. In *Conceptual Structures: Integration and Interfaces: 10th International Conference on Conceptual Structures,* ICCS 2002 Borovets, Bulgaria, July 15–19, 2002 Proceedings 10 (pp. 92-106). Springer Berlin Heidelberg.

[18] Slimani, T., BenYaghlane, B., & Mellouli, K. (2007, March*). Une extension de mesure de similarité entre les concepts d'une ontologie*. In International conference on sciences of electronic, technologies of information and telecommunications (Vol. 69).

[19] David Sánchez, Montserrat Batet, David Isern, and Aida Valls. Ontology-based semantic similarity: A new feature-based approach. Expert systems with applications, 39(9) :7718–7728, 2012.

[20] Pirró, G., & Euzenat, J. (2010). *A feature and information theoretic framework for semantic similarity and relatedness*. In The Semantic Web–ISWC 2010: 9th International Semantic Web Conference, ISWC 2010, Shanghai, China, November 7-11, 2010, Revised Selected Papers, Part I 9 (pp. 615-630). Springer Berlin Heidelberg.

[21] Le Ngoc, Luyen, Marie-Hélène Abel, and Philippe Gouspillou. "Apport des ontologies pour le calcul de la similarité sémantique au sein d'un système de recommandation." *Ingénierie des Connaissances (Evènement affilié à PFIA'22 Plate-Forme Intelligence Artificielle)*. 2022.

[22] Thomas R. Gruber, « *A translation approach to portable ontologyspecifications* », *KnowledgeAcquisition*, vol. 5, n^{o} 2, juin1993, p. 199220 (DOI 10.1006/knac.1993.1008)

[23] P. Resnik. Semantic similarity in a taxonomy: An information-based measure and its application to problems of ambiguity in natural language. Journal of Artificial Intelligence Research, 11:95-130, 1999.

[24] Seco, N., T. Veale, et J. Hayes (2004). An intrinsic information content metric for semantic similarity in wordnet. In ECAI, Volume 16, pp. 1089.

[25] Zhou, Z., Y. Wang, et J. Gu (2008). A new model of information content for semantic similarity in wordnet. In Future Generation Communication and Networking Symposia, 2008. FGCNS'08. Second International Conference on, Volume 3, pp. 85–89. IEEE.

[26] Harispe, Sébastien, et al. "Utilisation des fonctions de croyance pour l'estimation du contenu informationnel des concepts d'une ontologie." *Rencontre francophone sur la Logique Floue et ses applications LFA*. 2015.

[27] Yousfi, Aola, Moulay Hafid El Yazidi, and Ahmed Zellou. "CSSM: A Context-Based Semantic Similarity Measure." *2020 IEEE 2nd International Conference on Electronics, Control, Optimization and Computer Science (ICECOCS)*. IEEE, 2020.

[28] Zhou, Qiaoli, Zhiwen Jiang, and Fengling Yang. "Sentences similarity based on deep structured semantic model and semantic role labeling." *2020 International Conference on Asian Language Processing (IALP)*. IEEE, 2020.

ISSN: 2356-5608